DOCUMENT RESUME

ED 197 234 CG 014 881

AUTHOR Hogan, Joyce C.
TITLE Physical Requirements of the Workplace: Research Considerations for Personnel Selection.
PUB DATE Sep 80
NOTE 33p.: Paper presented at the Annual Convention of the American Psychological Association (88th, Montreal, Quebec, Canada, September 1-5, 1980). Best copy available.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability Identification: Criterion Referenced Tests: *Employment Qualifications: *Human Factors Engineering: *Job Analysis: Job Performance: Muscular Strength: *Performance Factors: Performance Tests: *Personnel Selection: Physical Fitness: State of the Art Reviews: *Work Sample Tests

ABSTRACT

Personnel selection processes for physically demanding jobs often discriminate against women, minorities, and handicapped individuals. Developing and validating job-related performance tests requires the assessment of the physical demands of a job and of the same physical capabilities in individuals. Job analysis is a systematic assessment procedure that must meet reliability and validity requirements in order to determine the critical tasks of a job. Two job analysis methods designed specifically to assess the physical demands of work are Fleishman's Physical Abilities Analysis and the Physical Effort Index, which produce information necessary for the development of work sample tests and simulations. These tests are currently being used and justified by content validity including single simulations of job content, integrated simulations, and criterion-referenced content valid tests. If factors important for performance are known, and reference tests are specified, physical ability tests should be able to predict future performance. Findings from three criterion-related validity studies suggest that future advances in physical ability testing will require input from several disciplines in order to respond to the need for effective personnel selection procedures. (NRB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

PHYSICAL REQUIREMENTS OF THE WORKPLACE: RESEARCH CONSIDERATIONS FOR PERSONNEL SELECTION

Joyce C. Hogan, Ph.D.

Center for Metropolitan Planning and Research
The Johns Hopkins University
Baltimore, Maryland 21218

CG 014881

Invited address to the Divisions of Industrial and Organizational Psychology and Rehabilitation Psychology, presented at the meeting of the American Psychological Association, Montreal, Canada, September 1980.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joyce C. Hogan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

This paper considers issues regarding developing and evaluating preemployment selection procedures for physically demanding jobs. Methods for assessing physical requirements in job performance provide indices of reliability, validity, and criticality of tasks within the overall job. The problem of translating critical task requirements into assessments of applicants is still unresolved. These job requirements can be tested primarily by means of work samples or simulations and physical ability tests. These testing alternatives are discussed and evaluated in terms of male and female subgroup results. Further consideration of these issues points toward the need to expand the conceptual model of physical abilities.

Personnel assessment for physically demanding jobs aims to identify applicants who can perform the work successfully. The defensibility of these kinds of selection processes is controversial because many physical tests discriminate, either fairly or unfairly, against women, handicapped individuals, and some minorities. Height and weight requirements, perhaps the most widely used physical standards for selection, are being thoroughly scrutinized for job relatedness. This has produced a tendency in employers simply to abandon selection requirements: nonetheless, they pull their heads out of the sand when productivity declines and worker compensation costs rise. We now see new interest in developing selection procedures for physically demanding jobs in the form of job related performance tests. Given that many physical tests result in adverse impact, the prudent employer will attempt to validate these tests. This paper describes approaches one might take to validate physical tests.

Selection procedures used to evaluate applicants for physically demanding work must meet the requirements specified in the Uniform Guidelines for Employee Selection (U.S. Equal Employment Opportunity Commission, 1978); in addition they must be consistent with the Section 503 or 504 provisions of the Vocational Rehabilitation Act of 1973 (U.S. Department of Labor, 1974). Minimum conditions for demonstration of validity are

specified in the Uniform Guidelines. These guidelines are perfectly consistent with professional standards of test construction and validation. Despite the existence of these formal test construction standards, the problems of developing and validating tests for selection in physically demanding jobs has yet to be definitively resolved. This paper is organized around two major issues - assessment of the physical demands of jobs and assessment of the same physical capacities in individuals. I would like to analyze each of these problems in more detail and suggest some ways in which they can be treated.

Assessments of the Physical Demands of Jobs

Assessing physical requirements of jobs is identical to, in its own way, job analysis. Although important steps in establishing rigorous job analysis methodologies have been provided by Christal (1974), Fine and Wiley (1971), Flanagan (1954), McCormick, Jeanneret, and Mecham (1972), Primoff (1971, 1972), and others, none specifically deal with physical work requirements. Moreover in conducting job analyses, the analyst often becomes overwhelmed by the sheer complexity of the subject matter that he or she is satisfied with an accurate inventory of the job content. Because of the biased nature of the information one often receives from incumbents in traditionally all male jobs, to rely only on an inventory of tasks would be ill-advised.

Job analysis as it is currently practiced is a quasi-intuitive art form. Sparks (1979) reviews considerations for the conduct of job analysis under the Uniform Guidelines which need not be repeated here. However, it is necessary to point out that job analysis is an assessment procedure and as such should attend to requirements of reliability, validity, and criticality. Using job analysis for assessing physical demands, quantitative estimates of reliability are provided by intraclass correlations of scale ratings made by job incumbents and supervisors. Ratings are typically scale scores from evaluations of tasks on performance dimensions. Qualitative estimates of reliability can be obtained through consensus of a panel of individuals (i.e., job analysts, subject matter experts) knowleageable about the job. Our confidence in either type of reliability is strengthened when incumbents and supervisors within a company, and analysts outside the organization all agree.

The validity of a job analysis is rarely questioned and even less frequently evaluated. Rater consensus regarding the task or set of tasks required for job performance is commonly but erroneously used to demonstrate validity. Strategies for validating a job analysis of physical labor include establishing on one hand retrospective or on the other empirical validity. Retrospective validity involves analyzing documents (i.e., work orders)

certifying work completed. From completed work, it is possible to document the tasks necessary to accomplish the work. Empirical validity demonstrates a relationship between a scale rating of a task and an objective measure of the same rating dimension. For example, if the metabolic costs expended to moor a ship are known, the validity of scale ratings of physical effort involved in the task can be ascertained.

Specification of the tasks that are critical for job performance is a major requirement of selection procedures that are supported by content or construct validity research. I refer to this as the problem of criticality and the problem exists because in the field, critical tasks often exist only in the eye of the beholder. Take for example the task in police work requiring that the officer support his partner while hanging from a ledge. Take also the task of transporting filled 55 gallon drums. It is often alledged that they must be manhandled, but the inquisitive job analyst will soon discover there is a specific technique for moving these drums.

Assessing abilities or tasks that are critical to job performance can be accomplished any number of ways. A model developed for several physically demanding jobs defines criticality in terms of ratings for scales of physical effort, frequency of performance, and importance of task performance (Hogan, Jennings, Ogden, and Fleishman, 1980). The rating means of each task are

then entered into a decision diagram and compared with a previously derived cutoff score. The result is a finite set of tasks that supersede in criticality all other physically demanding tasks of the job.

Determining the most critical tasks of a physically demanding job has a number of obvious advantages. First, a reduced set of tasks gives direction to test development and allows the developer to focus on certain key job requirements. On the other hand, the critical tasks can also aid criterion specification and evaluation. For example, ratings of performance on these tasks define the critical aspects of the job. Moreover, it is the critical tasks that should form the basis of the risk assessment in the preemployment medical examination. This is prerequisite for ensuring the job relatedness of the medical examination. Finally, if preemployment tests, performance appraisals, and medical examinations are rooted in critical tasks for job performance, handicapped applicants can be considered for employment based on critical, not ancillary or unimportant components of the job.

Given the foregoing considerations, two job analysis methods designed specifically to assess physical demands of work can be discussed. Perhaps the best known of these is Fleishman's (1976) Physical Abilities Analysis method. This uses nine factor analytically derived constructs based on the intercorrelations

of 60 athletic tests (Fleishman, 1964). Each factor is defined, distinguished from other factors, and used in combination with a seven-point, behaviorally anchored rating scale. Subjects who may be incumbents, supervisors, or job analysts use each scale to estimate the degree to which that physical ability is required for performance. The level of analysis is either by task or job; the result is a profile across the nine physical proficiency factors. Reliability of the method, defined in terms of intraclass correlations, routinely exceeds .7 with 20 raters. The validity of the scale scores is difficult to determine since no validity data are available. Therefore, predicting measures of ability requirements from scale scores based on this method is not yet possible.

A second technique for analyzing physically demanding jobs is called the Physical Effort Index (Hogan and Fleishman, 1979) and is a more general method than that of the Physical Abilities Analysis. Fleishman evaluates performance along nine separate dimensions, while this method uses only one -- generalized physical demands. This method is based on the psychophysical concept that individuals can use their perceptions of effort to estimate the physical requirements of a task. This is a scaling procedure where physical effort is defined for the rater and descriptors of high and low physical effort are provided. The seven point rating scale is anchored with examples whose physical costs are

known. The rater then uses the scale to indicate how much physical effort is necessary for performance of a job task. Although one could assess physical effort data for overall job performance, the typical procedure is to gather ratings on individual tasks listed for a job. The reliability of this method defined in terms of intraclass correlations averages about .8. Two studies have established the validity of this method for predicting and assessing metabolic and ergonomic costs of task performance (Hogan, Ogden, Gebhardt, and Fleishman, in press).

The foregoing discussion of assessing the physical demands of jobs can be summarized with the following points. First, job analysis is an assessment procedure that ought to take seriously requirements for reliability and validity. Second, critical tasks of a job need to be determined through systematic procedures. This is not only required legally, but has great practical importance. Finally, there are two job analysis methods that exist solely for the purpose of assessing physical requirements of jobs. Each is unique and each has considerable research support.

In the next section I would like to discuss assessing physical capacities of individuals to meet job requirements. This will focus on test development and the effectiveness of these tests in the selection process.

Assessments of Physical Capacities of Individuals for Work

Dunnette (1976, pp. 473-520), Fleishman, 1975, and others point out the difficulties of translating the requirements of jobs into the capacities of applicants. Taking this to the next logical step, even more difficulty is encountered in developing tests of those capacities. The problem is particularly acute when dealing with applicants for physically demanding work because the available tests of physical capacities are designed to assess athletic proficiency. Therefore, these tests may not necessarily be well suited for assessing the physical capacities required in occupational performance. There is a tendency to choose predictors out of expediency; if this is done, one should find validity evidence for the predictor prior to incorporation into a battery. Beyond a host of athletic tests, few occupationally derived physical performance tests exist and of those available, many tend to be content referenced for a specific job. The generalizability of these tests is, therefore, limited.

The kind of tests that an employer chooses to validate depends, in part, on the purposes and goals of selection as well as the nature of the information yielded from the job analysis. Tests for employment in physically demanding entry level jobs consist largely of ability measures and work samples (Hogan, 1980, pp. 75-78). The reasoning underlying predictor choice might be the validation strategy to be used. However, the categories of validity are not

as distinct as they might be with criterion-related content, and construct strategies often overlapping in attempts to demonstrate the validity of physical tests for work performance.

Development and Effectiveness of Work Sample Tests of Physical Performance

Job analytic procedures that identify the critical tasks necessary for job performance produce information necessary for the development of work sample tests or simulations. The use of these tests for employment is justified initially on the basis of content validity. Ideally, follow-up research will demonstrate predictive validity. Physical tests that are justified on the basis of content validity and are being used currently tend to fall into three categories: single simulations of job content, integrated simulations of job content, and criterion-referenced content valid tests. Each of these will be discussed in turn.

Single simulations of job content are usually developed directly from individual critical tasks. The number of independent critical behaviors is ascertained and single simulations of each behavior is designed. Either a discrete or a continuous scoring system is devised, preferably a continuous one where applicants can be ranked in terms of performance. Usually several of these single simulations are compiled as subtests into an overall battery, which provides more compre-

hensive coverage of the total job content than any single simulation. Perhaps the most visible recent example of single simulation physical testing is the physical performance test battery used for selection by the New York State Police (U.S. v State of New York, 1979). The first phase of the battery, which was scored discretely with pass/fail, consisted of tests entitled "Shotgun Aiming," "Physical Coordination Course," "Tire Change," "Portable Scale Lift" and "Deer Lift Off Roadway." The titles indicate the nature of the simulations. Phase II, for those individuals passing Phase I, consisted of subtests entitled "Policy Foot Pursuit Course," "Drag Person From Vehicle," "Attic Opening," and "Mile Run."

A critique of the tests used in this specific case is beyond the scope of this paper, but the results produced by these tests are interesting. Of the 3,706 applicants, 98.9% of the males passed, while 67.2% of the females passed indicating a selection rate for women less than 4/5th the rate for men. The judge ruled that certain parts of the physical performance test could have been scored differently serving the purpose of reducing the adverse impact on females at the same time maintaining the quality of the applicant pool.

The second category of content valid tests are integrated simulations. These tests attempt to combine critical job behaviors in a single simulation while maintaining fidelity of the work content. Integrated simulations are an efficient means for

evaluating several critical performances, they tend to retain the ecological validity of the workplace, and they constitute a reasonable job preview. A test developed recently and currently in use about which some effectiveness data are available exemplifies an integrated simulation.

An integrated simulation for grocery order selectors for a large wholesale grocery warehouse was developed based on the configuration of the warehouse and the volume of groceries moved by warehouse personnel over a period of one year (Hogan, Ogden, and Fleishman, 1979). The job analysis consisted of a detailed investigation of weight, physical dimensions, warehouse location, and sales rate for every grocery product in the warehouse. From this, a simulation could be constructed that represented a microcosm of the actual warehouse and required the same physical performances as those required on the job. The test involved locating, lifting, carrying, and loading 19 specified grocery cases as quickly as possible. Note that any one of these behaviors could be tested individually. The scoring system for the test was elapsed time from start to finish. Although this test could be supported by content validity, it was evaluated by a criterion related strategy where the criterion measure was an objective production index - cases moved per hour on the job. The validity coefficient for this test was .56 and an assessment of test fairness indicated no significant sex differences in terms of regression line slopes

or intercepts.

The third category of content valid tests are criterion - referenced content tests. These tests are designed to elicit performance that can be evaluated against an objective measure of the critical task. For example, if it is determined that tilting a 55 gallon drum filled with a chemical compound while using proper technique requires 80 pounds of force, a simulation measuring such a performance can be designed to indicate whether an applicant can exert that amount of force. The critical score is specified in advance; scores greater than the critical score will not reflect proportionately superior performance, because the ceiling has been reached (See American Psychological Association, Division of Industrial-Organizational Psychology, 1980). Consider, for example, one systematically developed test in this area is for entry level process operator jobs in a petroleum refining company (Osborne, 1977). The level of analysis was physical demands of operating valves during emergency situations; the job analysis indicated that this task was in fact critical for process operators in refineries. Emergency valves located throughout the refinery were sampled and the force and foot pounds of work required to "crack," "open" and "close" the valves were determined empirically. A simulated valve turning task was designed to measure the force produced by applicants; they could then be evaluated relative to the measurements taken in the field. The critical score consisted

of a specified number of valve turning repetitions, this score insured an appropriate level of physical strength in new employees, and 58% of the female applicants met the critical score.

It seems clear that, in the physical performance area, effective, content valid tests originate from a good job analysis. A job analysis that defines critical tasks representing valid elements of the job will provide some guidelines for test development. Content valid physical tests share some of the same strengths and weaknesses as other tests justified on the basis of content validity. The two major weaknesses I see are that (1) content tests are situationally specific, which severely limits their generalizability and (2) performance on content valid tests only describes behavior sampled from that particular content domain. Performance on external variables or overall measures of job performance cannot be assessed. On the positive side, content valid tests appear to meet employers' hiring needs and goals while allowing them to maintain an acceptable level of productivity. These tests also provide an alternative to predictive validity studies; this is important because such studies simply are not feasible for many employers. Employer's demands for immediate selection procedures will probably never be met by personnel researchers; nonetheless, the process of developing and implementing content valid physical tests is typically shorter than conducting a predictive validity study.

Finally, content valid tests provide a reasonable job preview for those who may be unfamiliar with the actual requirements of the work. This might be construed as having a chilling effect on applicants, however, immediate turnover is equally undesirable.

Ability Tests as Predictors of Criterion Performance

If an employer's purposes and goals in personnel selection are best served by a criterion-related validity strategy, the predictor battery likely to be considered will, no doubt, be physical ability tests - analogous to the cognitive ability tests of Guilford (1967) for the intellectual domain. As in other domains, physical ability tests can serve as reference tests for factors hypothesized to underlie performance. Theoretically, if factors important for performance are known and reference tests are specified, one should be able to predict future performance based on test performance.

Fleishman's (1964) factor analyses of physical proficiency tests provides the only empirically derived model for thinking about the structure of physical performance. His job analysis procedure described earlier is used to evaluate the physical requirements of work in terms of such dimensions. Once a significant physical requirement is identified, the test of the underlying ability can be identified by going back to the original factor analysis and choosing a test with a substantial

loading on the factor considered important. This test becomes the reference test of that ability and can be entered into the predictor battery along with other required ability tests.

Three criterion related validity studies completed to date reflect the effectiveness of physical ability tests to predict measures of job or training performance. All of these studies are based on Fleishman's (1976) Physical Abilities Analysis of job requirements and represent the collective wisdom of the field regarding physical ability tests in criterion-related validation research. Rather than discuss the details of each study, the findings are integrated under the topics of predictors of strength, flexibility, balance, and stamina.

Physically demanding work connotes strength requirements and when we think of physical testing, we tend to think of strength assessment. Our three studies focus more heavily on the strength predictors than predictors of other physical abilities. Jones and Prien (1978) report minimal overlap in the distribution of scores between 51 male and 10 female applicants placed in entry level, unspecified, physically demanding jobs. For the total group and the male subgroup, they report significant validity coefficients for a composite test score, which includes measures of static, explosive, and trunk strength, with a criterion of performance ratings. There is no report of subgroup validities for women and no reports of test fairness. With such a small

female sample, it would have been difficult to evaluate the effectiveness of these tests for predicting job success or the selection rates for women. Reilly, Zedeck, and Tenopyr (1979) included strength tests as one third of their predictor battery in studying the validity and fairness of physical ability tests for outdoor telephone craft jobs involving, among other tasks, ladder handling and pole climbing. These included two measures of static strength - grip strength and a cable pull across the chest - and three measures of dynamic strength - leg, trunk, and arm strength. Of these five strength measures, the cable pull across the chest enters the final predictor battery, and combined with scores on the two other tests produced a significant relationship between the total battery scores and time to complete training course. Comparisons between male and female regression equations indicated no significant differences, suggesting that the same equation was appropriate for selecting both males and females. Hogan, Ogden and Fleishman (1979) studied selection for entry level warehouse personnel with a predictor battery composed of one third strength tests. The Fleishman reference tests included were two static strength tests - dominant and non-dominant grip strength - test of dynamic strength - push-ups - and a test of trunk strength - leg lifts. Although all of these tests were valid for the total group where the criterion measure was an objective production index of units shipped per hour, none

were valid in either the male or female subgroup and none were entered into the final predictor battery. Table 1 summarizes the male and female subgroup validities of strength predictors of job or training performance in these studies. As seen these tests are not predicting performance for women and are only slightly more effective for men.

In the flexibility area, Fleishman distinguishes between extent and dynamic flexibility, with extent referring to the range of movement required in an activity, dynamic referring to the ability to make repeated, rapid flexing movements. Jones and Prien (1978) used the Fleishman reference test of dynamic flexibility that involves a cycle of twisting from floor to shoulder. In combination with the strength tests discussed previously, this test was part of the final predictor battery whose scores correlated with ratings of physical proficiency and job performance. Again, we have data on only the total group and male subgroup validity. No data on the female subgroup are provided. Reilly, Zedeck and Tenopyr (1979) included a single test of extent flexibility which involved measuring the distance the arm could be extended in a rotating stretch of the trunk. No significant validity coefficients were obtained for either total group or male and female subgroups. Hogan, Ogden, and Fleishman (1979) included a test of both extent and dynamic flexibility in the trial test battery to select warehouse personnel.

Dynamic flexibility was significantly correlated with the criterion measure for the total group but failed to result in subgroup validities for either males or females. The extent flexibility test did not correlate with the criterion and neither dynamic nor extent flexibility tests were entered into the final predictor battery. Table 2 summarizes these results and indicates that flexibility tests have not been predicting successful performance for either the male or female subgroup.

Reilly, Zedeck and Tenopyr (1979) found balance to be a significant requirement on the job and included in the predictor battery Fleishman's reference test for balance. This is a measure of the total time in 2 trials one can balance on a 3/4" beam with one foot. The test was a valid predictor of time to complete the training program for the total group as well as both male and female subgroups. It was entered as the second of three tests in the final test battery.

Stamina or cardiovascular endurance is a factor whose unofficial status is that of a stepchild in personnel research assessments. Predictive assessments of stamina tests are typically not carried out by the personnel specialist, because prudent test administration would require a physician to be present. A strong argument could be made for placing cardiovascular assessments into the medical domain. If these evaluations are made by physicians, they should be job-related and the risk assessment should be

validated in terms of epidemiologic or other types of medical evidence. If tests of stamina are to be administered by personnel specialists, the most appropriate form would be criterion-referenced content valid tests where once a critical score is obtained the test is terminated.

Although there is only a small amount of data available to evaluate physical ability testing, evidence regarding the effectiveness of these tests for predicting criterion performance is not overwhelming. One could point to insufficiencies in either the tests or the criterion measures, however in the studies discussed above, the criteria are fairly strong. This raises the question of the effectiveness of ability tests particularly in the strength area. This is far and away the greatest concern of employers and this is the area in which sex differences tend to be so vivid. The Hogan, Ogden and Fleishman study found the traditional strength measures to be biased since the criterion score predicted from the common regression line was consistently too high for men and too low for women. That is, women tended to do better on the job than would have been predicted from the test scores. Strength is such an important dimension in physically demanding work that we have to develop better measures that correspond to work, not athletic performance.

Discussion

I would agree with Guion (1976, pp. 777-828) that one of the most serious shortcomings in selection research is the limited generality of most findings. In physical testing, findings are often more situationally bound because we have just begun developing technically adequate content valid tests. Our physical ability tests are not predicting job performance as powerfully as we would like. Yet, it is clearly these kinds of tests that can move us beyond situationalism and a limited technology. Can we deal with the problem of generalization in physical testing? Consider some reflections about the basis for physical testing.

The model is factor analytically derived and seeks to determine dimensions of physical performance. Conceptually, this approach has much merit. A limitation may be the content out of which these dimensions are interpreted. The physical abilities content was originally intended to focus on assessments of athletic proficiency and this content seems limited for work in occupational settings. There are four problems that come to mind. First, strength assessments in the athletic domain measure individual muscle groups. However, people working use a variety of muscle groups that together result in a single force. Perhaps more appropriate tests would capitalize on this distinction.

Second, there are some critical work behaviors that have no analog in the athletic ability testing sphere. Two examples are

(1) manual materials handling movements such as pushing, pulling, carrying, and lowering and (2) movements involving torque such as cranking and turning.

Third, athletic tests do not take account for the technique or capacity to acquire technique which facilitates physical performance. This whole area of neuromuscular coordination is relatively unexplored as it applies to industrial performance. There are case reports of physically unimpressive women performing heavy work successfully simply because they adopted biomechanically proper and advantageous techniques.

Fourth, athletic tests assess high levels of an individual's physical proficiency whereas tests appropriate for occupational physical performance should be concerned with the minimal requirements for effective performance. This is particularly important for setting the critical score in performing of criterion referenced content valid tests.

There are no immediate solutions. But it is clear that the present system of ad hoc test development will not meet the needs of scientific generalizability. An approach that does seem fruitful is the use of a construct validation method. In order to even apply this strategy, one must take seriously the structure of physical abilities required in occupational task performance. This structure requires that the important performance categories be

specified a priori - not constructed on a purely empirical basis. These performance categories should meet at least 4 preliminary requirements. First, they should adequately map the domain of physical abilities. This insures generalizability of the emergent tests. Second, these performance categories should be trainable. This has the effect of reducing adverse impact. Third, the categories should be concepts that permit variability on both criterion and predictor measures. This enhances the statistical possibility of validity. And fourth, the content of the performance categories should be face valid. This will result in fewer applicant challenges of the test's legitimacy.

Finally, we have to begin at the conceptual level if we want generalizability. This is not the same as beginning at the beginning because we have the benefit of research carried out by a handful of pioneers such as Fleishman. Future advances in physical ability testing will require input from several disciplines including work physiology, biomechanics, ergonomics, as well as industrial psychology. Research in this area will not only constitute an advance within each of these involved disciplines, but will respond to a major practical problem in employment selection.

References

American Psychological Association, Division of Industrial - Organizational Psychology. Principles for the validation and use of personnel selection procedures (2nd ed.) Berkeley, CA: Author, 1980.

Christal, R. E. The United States Air Force occupational research Project. (Technical Report), Brooks Air Force Base, Texas: Air Force Human Resources Laboratory, 1974.

Dunnette, M. D. Aptitudes, abilities, and skills. In M. D. Dunnette (d.), Handbook of industrial and organizational Psychology. Chicago: Rand McNally, 1976.

Fine, S. A. and Wiley, W. W. An introduction to functional job analysis: Methods for manpower analysis (Monograph No. 4). Kalamazoo, Michigan: W. E. Upjohn Institute, 1971.

Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.

Fleishman, E. A. The structure and measurement of physical fitness. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Fleishman, E. A. Toward a taxonomy of human performance. American Psychologist, 1975, 30, 1127-1149.

Fleishman, E. A. Physical abilities analysis manual. Silver Spring, MD: Advanced Research Resources Organization, 1976.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Guion, R. M. Recruiting, Selection, and Job Placement. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Hogan, J. C. The state of the art of strength testing. In D. C. Wash and R. H. Egdahl (Eds.). Women, work, and health: challenges to corporate policy. New York: Springer Verlag, 1980.

Hogan, J. C., and Fleishman, E. A. An index of physical effort required in human task performance. Journal of Applied Psychology, 1979, 64, 197-204.

Hogan, J.C., Ogden, G. D., Fleishman, E. A. The development and validation of tests for the order selector job at Certified Grocers of California, Ltd. (Technical Report). Washington, D.C.: Advanced Research Resources Organization, 1979.

Hogan, J. C.; Jennings, M. C.; Ogden, G. O.; and Fleishman, E. A. Determining the physical requirements of Exxon apprentice jobs. Washington, D.C.: Advanced Research Resources Organization, 1980.

Hogan, J. C.; Ogden, G. O.; Gebhardt, D. L., and Fleishman, E. A. Reliability and validity of methods for evaluating perceived physical effort. Journal of Applied Psychology, in press.

Jones, M. A. and Prien, E. P. A valid procedure for testing the physical abilities of job applicants. Personnel Administrator, 1978, 23, 33-38.

McCormick, E. J.; Jeanneret, P. R., and Mecham, R. C. A study of job characteristics and job dimensions as based on the position analysis questionnaire (PAQ). Journal of Applied Psychology, 1972, 56, 347-368.

Osburn, H. G. An investigation of applicant physical qualifications in relation to operator tasks at the Deer Park Manufacturing complex. (Technical Report). Houston, Texas: Employment Services, Head Office, Employee Relations, Shell Oil Company, 1977.

Primoff, E. S. Summary of job-element principles: Preparing a job-element standard. Washington, D.C.: Personnel Measurement and Development Center, U. S. Civil Service Commission, 1971.

Primoff, E. S. The J-coefficient procedure. Washington, D.C.: Personnel Measurement and Development Center, U.S. Civil Service Commission, 1972.

Reilly, R. R.; Zedeck, S.; and Tenopyr, M. L. Validity and fairness of physical ability tests for predicting performance in craft jobs. Journal of Applied Psychology, 1979, 64, 262-274.

Sparks, C. P. Job analysis under the new Uniform Guidelines (Technical Report). Houston, Texas: Personnel Research,

Employee Relations Department, Exxon Company, U.S.A., 1979.

U.S. Department of Labor, Employment Standards Administration, Employment of the Handicapped. Affirmative action obligations of contractors and subcontractors. Federal Register, 1974, 39, 20566-20571.

U.S. Equal Employment Opportunity Commission, U.S. Civil Service Commission, U.S. Department of Labor, and U.S. Department of Justice. Adoption by four agencies of uniform guidelines on employee selection procedures (1978). Federal Register, 1978, 43, 38290-38315.

U.S. v State of New York. Civil Action No. 77-CV-343. U.S. District Court, Northern District of New York, 475 F. Supp. 1103-1110.

TABLE 1

VALIDITY OF STRENGTH PREDICTORS
FOR JOB OR TRAINING PERFORMANCE

STUDY	TEST	STATIC ♂	STATIC ♀	EXPLOSIVE ♂	EXPLOSIVE ♀	DYNAMIC ♂	DYNAMIC ♀	TRUNK ♂	TRUNK ♀
Jones & Prien (1978)	composite score [a,b]	*		*				*	
Reilly, Zedeck, & Tenopyr (1979)	static strength	.20*	.06						
	grip strength	-.25	-.14						
	dynamic arm					.06	-.12		
	dynamic leg					-.07	-.12		
	dynamic trunk					-.15	-.18		
Hogan, Ogden & Fleishman (1979)	grip strength - dominant	.09	-.11						
	grip strength - non-dominant	.14	.15						
	push-ups					-.03	-.33		
	leg lifts							-.07	-.33

a - no individual test score correlations reported
b - no female subgroup validites reported
* - $p < .05$

TABLE 2

VALIDITY OF FLEXIBILITY PREDICTORS

FOR JOB OR TRAINING PERFORMANCE

STUDY	TEST	EXTENT ♂	EXTENT ♀	DYNAMIC ♂	DYNAMIC ♀
Jones & Prien (1978)	composite score[a,b]			*	
Reilly, Zedeck, & Tenopyr (1979)	extent flexibility	-.31	-.22		
Hogan, Ogden, & Fleishman (1979)	sit and reach	.03	.15		
	dynamic flexibility			.17	.04

a - no individual test score correlations reported
b - no female subgroup validities reported
* - $p < .05$